

"Express Mail" mailing label number:

EL764880479 US

CLASSIFICATION ENGINE FOR MANAGING ATTRIBUTE-BASED DATA

Ben Folk-Williams
Jimmy Wan

BACKGROUND OF THE INVENTION

5 **Field of the Invention**

The present invention generally relates to computer systems and computer databases, and more specifically to a software engine that may be used by various application programs as the primary interface for viewing, searching, and modifying information stored in databases, particularly relational databases which are used to
10 maintain electronic catalogs.

Description of the Related Art

Computer systems use databases to maintain information for a wide variety of purposes. A database may generally be thought of as a file composed of one or more
15 records, each containing fields, together with a set of operations used for searching, sorting, recombining, and other functions. Data in a typical relational database can be represented as a set of records (entries) spread out across multiple tables. Entries in a table are comprised of values for several data columns. The database has a definition for each table which dictates the format of records in the database, including the
20 number of fields, specifications regarding the type of data that can be entered in each field (e.g., numeric or alphabetical), and the field names used. A database is physically stored on any conventional computer storage medium, such as a hard disk drive. A database server (a network node or workstation) may be dedicated to storing and providing access to a shared database.

Applications can be custom-written to access a database having a particular schema (the format of the database). Alternatively, a database engine (a program module or modules) is often used to provide access to a database management system. The database management system (DBMS) is the software interface between the database and the user. A database management system (or database manager) handles user requests for database actions, and allows for control of security and data integrity requirements.

One important use for databases is in electronic catalogs. With the explosive growth in the number of users of the Internet, particularly the World Wide Web (WWW), many retailers are taking advantage of the opportunity to present goods and services to consumers via this new purchasing medium (e-commerce). Online catalogs are often represented as a hierarchy or tree wherein each level of the hierarchy represents a slightly more specific classification about particular products available in the catalog. For example, products from an automobile dealership might first be categorized according to the particular vendor (manufacturer), then according to a particular body style, and then according to a particular color. Information regarding all of the available cars/trucks accessed through a database by the web server or application server, and updated periodically. Users are able to search the catalog using limited search criteria based on the available fields.

As suggested above, and as illustrated in Figure 1, a web application 2 may be customized to directly access a catalog database 4. In this situation, the web application is programmed based on the knowledge of the specific database structure of the catalog database, i.e., the number of total database files in the overall database 4, and the fields in each of the database files. Web application 2 can include a search function that formulates a request based on user inputs, and generates a query to catalog database 4 according to a static list of attributes.

As further suggested above, and as illustrated in Figure 2, a web application 6 may be designed to access a catalog database 8 via a catalog engine 10. In this situation, web application 6 need not be programmed according to the specific database structure of catalog database 8. Rather, web application 6 is only required to interface with catalog engine 10, which is then responsible for interacting with the

database. For example, a conventional catalog engine might be used to carry out search operations by generating a SQL query. A SQL query uses structured query language, a database language that has become the standard for most database products. Catalog engine 10 may be used by other applications as well, such as
 5 another web application 12 (or a Java applet, etc.).

One major problem with these electronic catalog systems relates to the fixed manner in which the application, or the catalog engine, interfaces with the catalog database. The structure (schema) of the database may need to be changed, in order to optimize it in a particular way, or to add new fields and capabilities. For example,
 10 Figure 3 is a representation of the conversion of information pertaining to an automobile dealership's inventory, from a first schema with one table 14, to a second schema with three tables 16, 18 and 20. In the first schema, the single table 14 contains all the information (that is, all fields) pertaining to each record (car/truck). In this simplified example, a given record has seven fields: Year (of manufacture); Mfr.
 15 (manufacturer); Doors (the number of doors); type (the automobile body design); model (a brand name); color; and VIN (vehicle identification number).

It can be seen that the structure of the single table 14 lends itself to certain inefficiencies. While sufficient when considering a single record, it becomes extremely inefficient when looking at all of the data as an aggregate. For example,
 20 much of the stored data is redundant. The first six entries of table 14 show how the corresponding six cars are all year 2000 Ford 4-doors. The first five entries are furthermore all Taurus models. It is possible to reduce this inefficiency by creating multiple tables which are interrelated, and which effectively reuse the redundant information. In the second schema, a Base table 16 is used to collectively identify all
 25 vehicles according to their year, manufacturer, and model. A Base ID (identification) value is then assigned to each record. This Base ID is used to correlate those records with entries in a Style table 18. In this example, the first four rows of Style table 18 all have a Base ID of zero which, according to Base table 16, corresponds to any vehicle that is a year 2000 Ford Taurus. Thus, the information that pertains to the
 30 fields in Base table 16 does not have to be repeatedly included in Style table 18, reducing the redundancy. Style table 18 adds information pertaining to the number of doors, vehicle type, and color, and then assigns a Style ID to each record. The Style

ID is then used to access the corresponding VIN in VIN table 20. This correlation between the tables of the second schema is the essence of a relational database.

The static nature of prior art catalog applications and catalog engines makes it difficult to accommodate such changes in a database. In most cases, it will be necessary to rewrite the application or engine, which can be extremely laborious (or impossible for the end user that may not have access to the application's original code). Some prior art catalog engines can create different classifiers for different attributes, but these classifiers must conform to a fixed database format, and changes nearly always necessitate extensive import/export operations. These problems are particularly troublesome for volatile databases (whose structure might change often). It would, therefore, be desirable to devise a more flexible catalog engine that can easily support changes in the catalog database schema. It would be further advantageous if the improved catalog engine were highly scalable for very large database applications.

SUMMARY OF THE INVENTION

It is therefore one object of the present invention to provide an improved classification system.

It is another object of the present invention to provide such an improved classification system that easily accommodates changes in the underlying database schema.

It is yet another object of the present invention to provide an improved method of constructing an electronic catalog using such a classification system having schema independence, which imparts performance increases when the database is optimized for specific situations.

The foregoing objects are achieved in a method of managing a computer database (e.g., a database for an electronic catalog) generally comprising the steps of importing data into a database residing on a computer system, constructing a schema object to represent the database schema, and using an aggregate classifier to

manipulate the database. The schema object may be constructed by defining a plurality of classifier definitions corresponding to specific columns and tables in the database. The classifiers may include: a "property" classifier which interacts with a single column on a single table; an "object" classifier which contains one or more of the "property" classifiers; a "split-object" classifier which makes more than one "object" classifier appear as a single classifier; a "join" classifier which identifies how multiple database objects are linked in a "split-object" classifier; and a "mapped property" classifier as a special form of the "split-object" classifier to manage data stored in a table of the database which serves as an index to another database table.

Parameterized classifiers may also be defined which are templates for classifiers that are instantiated when associated parameters are provided.

The invention allows the classification system to modify the database structure and easily conform the classification engine to the modified structure without recompiling the engine or rewriting the catalog application. The engine is conformed to the new structure by constructing a second schema object for the modified database. The schema objects are preferably constructed from classifier definitions that are defined using a field-based language such as extensible markup language (XML).

The above as well as additional objectives, features, and advantages of the present invention will become apparent in the following detailed written description.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention, its numerous objects, features, and advantages may be better understood, and made apparent to those skilled in the art by referencing the accompanying drawings.

FIG. 1 is a block diagram of a simple prior art electronic catalog system, wherein a static web application directly manages a catalog database;

FIG. 2 is a block diagram of another prior art electronic catalog system, wherein a static catalog engine provides an interface between the web application and the catalog database;

FIG. 3 is a representation of the conversion of an exemplary catalog database
5 from an old single-table schema to a new 3-table schema;

FIG. 4 is a block diagram of one implementation of an electronic catalog system constructed in accordance with the present invention, having a dynamic classification engine which easily adapts to different schemas of the underlying catalog database; and

10 FIG. 5 is a block diagram of one embodiment of the classification engine of the present invention, which uses an XML (extensible markup language) document, representing the database schema, to construct a schema object.

The use of the same reference symbols in different drawings indicates similar or identical items.

15

DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

With reference now to the figures, and in particular with reference to Figure 4, there is depicted one embodiment 30 of an electronic database system constructed in accordance with the present invention. Electronic database system 30 is generally
20 comprised of a catalog database 32, a classification engine 34, several applications including a web application 36, swing tools 38, or other applications 40, and a content management tool 42. While the invention is described in the useful context of electronic catalogs, those skilled in the art will appreciate that it is not so limited, and will understand after reading the following description that the invention is applicable
25 to database management in general.

In the illustrative implementation, database system 30 is for adapted for e-commerce, and runs on a conventional network server which is connected to, e.g., the Internet. Thus, database system 30 is physically embodied in the program instructions

and data that reside on the network media, that is, storage media (local or remote) or transmission media. The network server includes a communications device, such as an Ethernet card, one or more processing units, and system memory that temporarily stores the program instructions to be executed.

5 Applications 36, 38 and 40 access catalog database 32 via classification engine 34. Classification engine 34 can classify any set of information into various groups by examining different values of common properties. For example, the vehicles of the tables of Figure 3 have attributes (classifiers) such as year, model, etc. In the present invention, classification engine 34 is dynamically modifiable to support
10 different schemas of catalog database 32 using any desired attributes and table structures. Classification engine 34 enables this functionality by constructing a schema object which defines both the classifiers available to the engine, and the mapping between these classifiers and objects in catalog database 32.

As further illustrated in Figure 5, classification engine 34 dynamically
15 constructs the schema object 44 using database schema definitions 46 provided via an extensible markup language (XML) document. XML is a field-based language similar to the hypertext markup language (HTML). Both of these languages provide a protocol for transmitting formatted information and control codes. Different fields within the XML document 46 are used to define the database structure and attributes,
20 using tags. In XML, a tag is a pair of angle brackets (<>) that contain one or more letters and numbers between the angle brackets. One pair of angle brackets is often placed before an element, and another pair placed after, to indicate where the element begins and ends. For example, the definition of a table structure might begin with the tag "<Table>", and end with the tag "</Table>".

25 The XML document instructs classification engine 34 how to map onto the database structure, and provides a convenient mechanism to easily change these mappings. Rewriting of the XML code may be understood with reference to the database conversion example of Figure 3. For the original schema, the XML code might appear as follows:

30 <Table name = OldCarSchema>

```

    <Property name = "Model Year"
        Column name = "Year"/>
    <Property name = "Manufacturer"
        Column name = "Mfr"/>

```

5 ...

 </Table>

This original code defines a single table (named "OldCarSchema") with the various attributes (only two of which are shown above, with the others indicated by the ellipses). For the second schema, the XML code might appear as follows:

```

10           <Composite_Object name = "NewCarSchema">           Composite_Object
           <Table name = Base>
               <Property name = "Model Year"
                   ...
           </Table>
15           <Table name = Style>
               <Property name = "Doors"
                   ...
           </Table>
           ...
20           </Composite_Object >

```

In this modified code, a composite object (named "NewCarSchema") has multiple tables (only two of which are shown above), and each table has defined attributes.

By defining the schema in XML, it is not necessary to recompile code when modifying the underlying database structure. XML document 46 need only be

25 updated so that it refers to the correct objects, and the application will then work without any further changes. To create a schema object, the XML document 46 is examined to define the set of classifier definitions. Classification engine 34 may provide many different standard classifier objects to handle a variety of different situations. Each kind of classifier is generally built to handle one kind of data. In the

illustrative embodiment, the most granular classifier is the "property" classifier, which maps to a single column on a single table in the database. Higher-level classifiers contain one or more lower-level classifiers. For example, the "object" classifier contains one or more "property" classifiers. Any classifier that contains other
 5 classifiers is referred to herein as an aggregate classifier.

Other standard classifiers built-in to classification engine 34 include: a "split-object" classifier; a "join" classifier; a "mapped property" classifier; and a generic aggregate classifier. The "split-object" classifier makes more than one database object appear as a single classifier, while a "join" classifier defines how multiple
 10 "object" classifiers should be linked in a "split-object" classifier. The "mapped property" classifiers are a special form of the "split-object" classifiers which are used to manage data stored in a table which serves as an index to another database table (e.g., the Base ID or Style ID fields in the second schema of Figure 3). Parameterized classifiers may also be used to generate classifiers on-the-fly, at run time.
 15 Parameterized classifiers utilize a template defined in XML, but they cannot be used in code until after the parameters have been provided. They are instantiated when the additional information is complete. Other classifiers may be designed by the end user, e.g., by extending the provided Java classes and adding customized functionality.

20 An XML document can also be used for importing data, i.e., modifying the contents (rather than structure) of catalog database 32. The import tool in classification engine 34 can import data into any kind of database that can be represented by an instance of database schema definitions 46. The import tool can be used even if the classification engine itself will not be used to view or manipulate this
 25 data once it has been imported into the database. The import tool determines the proper SQL based on the database schema definitions previously used in designing the database schema. The <Create/> tag can be used to add data. The <Delete/> tag can be used to remove records from the database. The <Modify/> tag can be used to manipulate records already in the database. The code that sets up the classifier values
 30 may also be nested to avoid redundant specification of related attributes on multiple records to be created, for example:

```

<!-- Specifies that the value 2000 is to be associated with the classifier "year" -->
<Classifier name="Year" value="2000">
  <Classifier name="Mfr" value="Ford">
    <Classifier name="Doors" value="4">
      <Classifier name="Type" value="Sedan">
        <Classifier name="Model" value="Taurus">
          <Classifier name="Color" value="White">
            <Classifier name="VIN" value
              = "WAT">
              <Create/>
            </Classifier>
          </Classifier>
        <Classifier name="Color" value="Red">
          <Classifier name="VIN" value="RAT">
            <Create/>
          </Classifier>
        </Classifier>
      </Classifier>
    </Classifier>
  </Classifier>
</Classifier>

```

Classification engine 34 also handles catalog searches handed down from the various web applications. The engine's interface to the application works with a top-level object (e.g., "Composite_Object"). The engine is able to work with an abstract set of attributes regardless of the location in the database schema. When an application uses classification engine 34, it passes a request to classification engine 34 to obtain an aggregate classifier based on a given schema object. The schema object creates an aggregate classifier based on its classifier definitions and returns it to the application. The application then passes a series of one or more requests to the aggregate classifier 48. Aggregate classifier 48 receives these requests and interrogates schema object 44 to find the locations of different classifiers. Schema object 44 returns the location information, e.g., that the "VIN" classifier is in the VIN table, the "Color" classifier is in the Style table, etc. Aggregate classifier 48 then takes the search constraints from the request, e.g., numdoors > 2, model = "Taurus", associates them with the relevant columns in the database tables, and passes this information to a query generator 50.

Query generator 50 converts the request to a SQL query. Conventional numeric, string, and Boolean operators may be used in the query. The engine generates the SQL query by comparing the user input with the schema, e.g.:

SELECT VINTable.VIN, Style.Color, Base.Year

5 FROM VINTable, Style, Base

WHERE VINTable.StyleID = Style.StyleID AND

Style.BaseID = Base.BaseID AND

Style.Doors>2 AND

Base.Model = "Taurus"

10 This query requests the vehicle identification numbers, colors, and years of manufacture for all cars that are of the model "Taurus" and have more than two doors (i.e., the 4-door sedans, and the 5-door station wagon).

A query executor 52 carries out the query by communicating directly with catalog database 32. Query executor 52 also processes the results from the database, and returns the processed data to aggregate classifier 48. Aggregate classifier in turn passes the results to the requesting application.

15

The present invention provides a flexible classification engine that may easily be adapted for use with any database schema, including that of a complete legacy database. Parametric searches can be performed across multiple attributes. All of this functionality is can be adapted to new database schemas without recompiling, by simply re-writing the XML document containing the classifier definitions. An application only has to interact with the classification engine, and need know nothing regarding the schema of the database. The engine also scales to large databases and large numbers of users with little or no difficulty, since most of the engine's operating data can be shared among different users of the system.

20

25

Although the invention has been described with reference to specific embodiments, this description is not meant to be construed in a limiting sense.

Various modifications of the disclosed embodiments, as well as alternative
embodiments of the invention, will become apparent to persons skilled in the art upon
reference to the description of the invention. It is therefore contemplated that such
modifications can be made without departing from the spirit or scope of the present
5 invention as defined in the appended claims.

FILED IN 2007